

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2022

Prof. Dr. Jens Stoye · Dr. Marília D. V. Braga

<https://gi.cebitec.uni-bielefeld.de/teaching/2022summer/sa>

Übungsblatt 4 vom 28.4.2022

Abgabe am 5.5.2022 bis 12:00 Uhr (mittags)

Aufgabe 1 (Eigenschaften von Alignments)

(3 Punkte)

1. Welche Zeichen dürfen in einer Alignmentsspalte bei gegebenem Alphabet Σ vorkommen? Welche Kombination ist nicht erlaubt?
2. Seien x und y zwei Sequenzen über Σ , A ein globales Alignment von x und y , und E eine Editsequenz, die A entspricht. Dann gilt:

$$\max\{|x|, |y|\} \leq |E| \leq |x| + |y|.$$

Konstruiere Extremfälle für die Ungleichheiten mit $\Sigma = \{0, 1\}$, $|x| = 1$ und $|y| = 2$.

Aufgabe 2 (Beziehung Edit-Sequenz–Globales Alignment)

(6 Punkte)

Gegeben seien die Sequenz $x = \text{ATGCTAG}$ und die Edit-Sequenz $E = \mathcal{I}_C \mathcal{C} \mathcal{C} \mathcal{S}_A \mathcal{D} \mathcal{C} \mathcal{C} \mathcal{C} \mathcal{I}_C$.

1. Gib das Alignment A an, welches der Edit-Sequenz E , angewandt auf x entspricht und berechne die Projektion $y = \pi_{\{2\}}(A)$.
2. Wie hoch sind die Einheits-Kosten der Edit-Sequenz E ?
3. Betrachte nun die Edit-Scores: $\mathcal{I} = \mathcal{D} = -\frac{1}{2}$, $\mathcal{S} = 0$ und $\mathcal{C} = 1$. Was ist dann der Score von E ?
4. Gib ein anderes globales Alignment der Sequenzen x und y an (kürzer als A).

Aufgabe 3 (Alle globalen Alignments)

(6 Punkte)

Gegeben seien die Sequenzen $x = \text{BA}$ und $y = \text{AB}$.

1. Zeichne den globalen Alignmentgraphen für x und y mit *allen* Kanten und Knoten. Füge deiner Zeichnung die Kantenbeschriftungen λ hinzu.
2. Gib alle möglichen globalen Alignments von x and y an.
3. Betrachte die folgenden Scores für die Editoperationen: $\mathcal{I} = \mathcal{D} = -\frac{3}{2}$, $\mathcal{C} = +1$ und $\mathcal{S} = -1$. Füge deiner Zeichnung die Kantengewichte w hinzu und berechne alle Knotenwerte $S(v)$.
4. Markiere in deiner Lösung zu 2. die optimalen Alignments und markiere deren Pfade im Graphen.

Aufgabe 4 (Optimaler Score)

(5 Punkte)

Die Rekurrenz zur Berechnung der Edit-Distanz mit Einheitskosten lautet

für $1 \leq i \leq x , 1 \leq j \leq y $: $D(i, j) = \min \begin{cases} D(i-1, j-1) + \mathbb{1}_{\{x[i] \neq y[j]\}} \\ D(i-1, j) + 1 \\ D(i, j-1) + 1 \end{cases}$	mit den Basisfällen $\begin{cases} D(0, 0) = 0, \\ D(i, 0) = i \text{ für } 1 \leq i \leq x \text{ und} \\ D(0, j) = j \text{ für } 1 \leq j \leq y \end{cases}$
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------

Betrachte nun die folgenden Scores für die Editoperationen:

Für Insertion und Deletion: $\mathcal{I} = \mathcal{D} = -2$.

Für Copy (\mathcal{C}) und Substitution (\mathcal{S}) gilt die Scorematrix

$\mathbb{M}(X, Y)$, für $X, Y \in \{A, C, G, T\}$:

M	A	C	G	T
A	+1	-2	-1	-2
C	-2	+1	-2	-1
G	-1	-2	+1	-2
T	-2	-1	-2	+1

1. Gib die Rekurrenz und ihre Basisfälle zur Berechnung des optimalen Edit-Scores von zwei Sequenzen x und y (über $\Sigma = \{A, C, G, T\}$) an.
2. Benutze die Rekurrenz zur Berechnung der Edit-Matrix und des optimalen Edit-Scores der Sequenzen $x = \text{TCTA}$ und $y = \text{TGCA}$.

Wie viele optimale Alignments/Edit-Sequenzen gibt es? Schreibe ein optimales Alignment auf.